TIB LEIBNIZ INFORMATION CENTRE FOR SCIENCE AND TECHNOLOGY UNIVERSITY LIBRARY

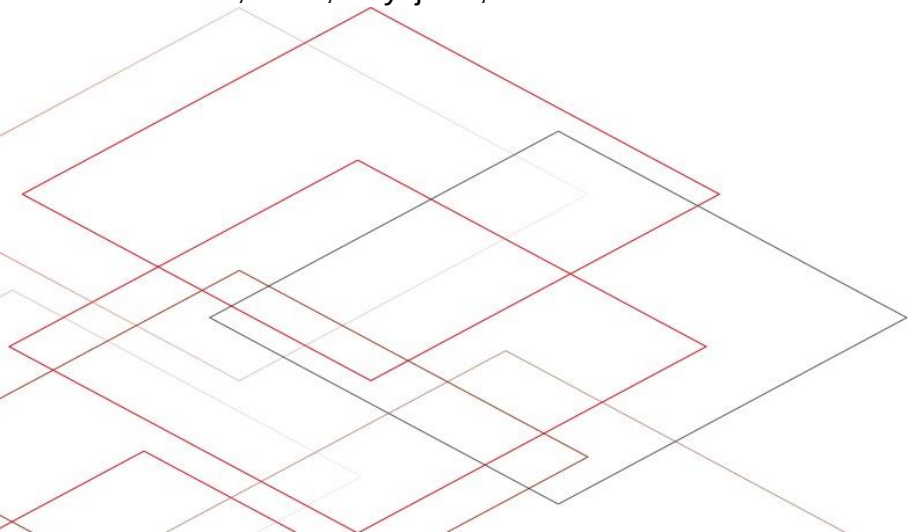
Introduction:
Open Research Knowledge Graph

Oliver Karras, Alessio Ferrari, Davide Fucci, and Davide Dell'Anna

oliver.karras@tib.eu, alessio.ferrari@isti.cnr.it, davide.fucci@bth.se, d.dellanna@uu.nl

32nd IEEE International Requirements Engineering 2024 Conference – Exploring New Horizons: Expanding the Frontiers of Requirements Engineering

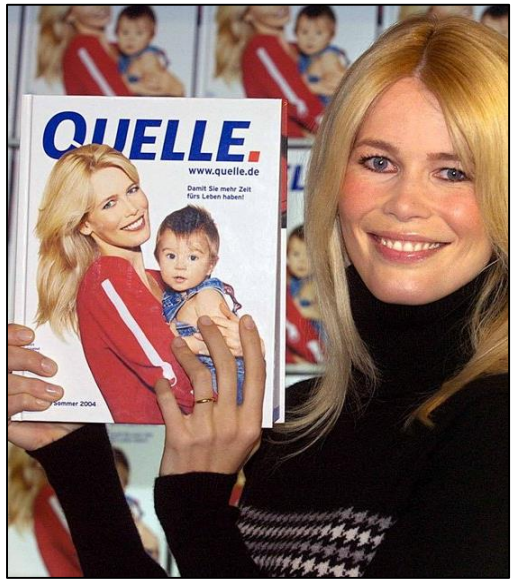
June 24th, 2024, Reykjavik, Iceland



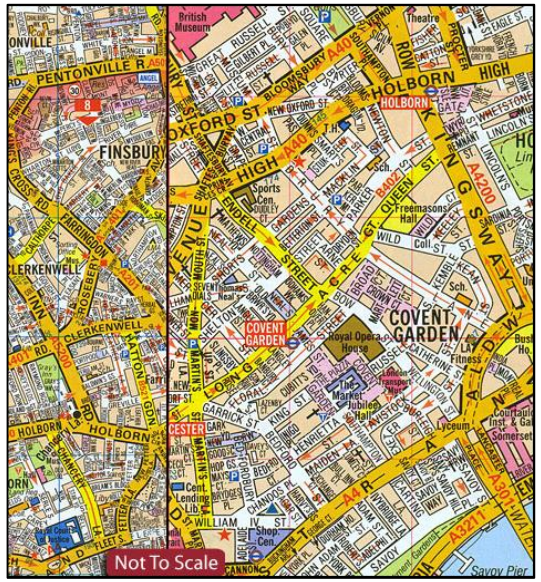
Leibniz Association

Once Upon a Time, we Communicated with Paper…

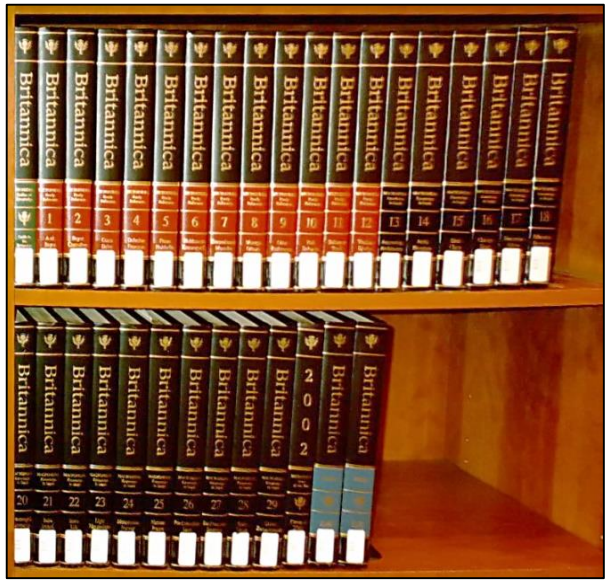
Who still remembers?



Mail order catalogs



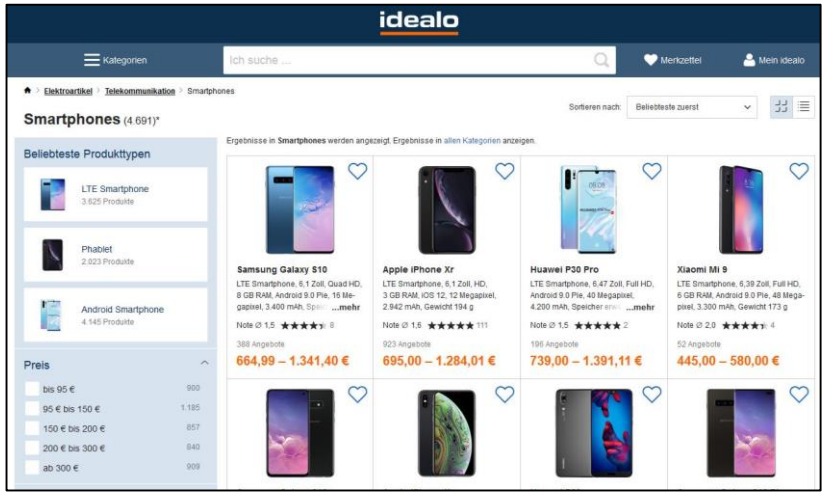
Maps

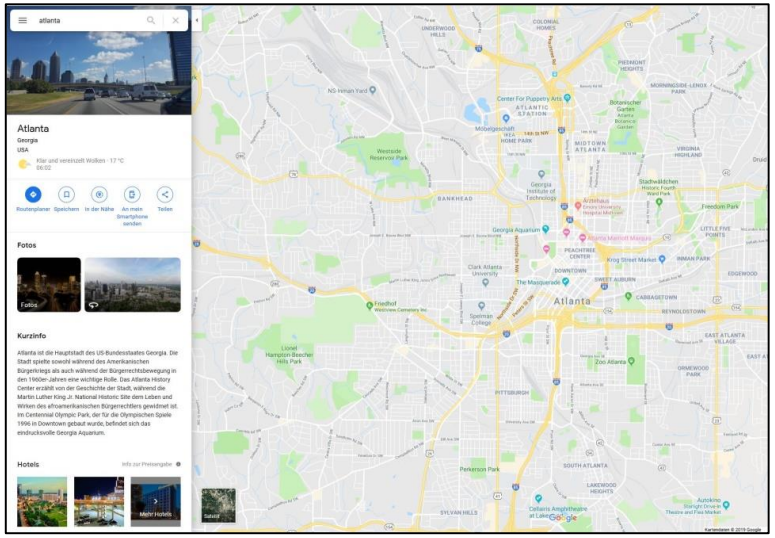


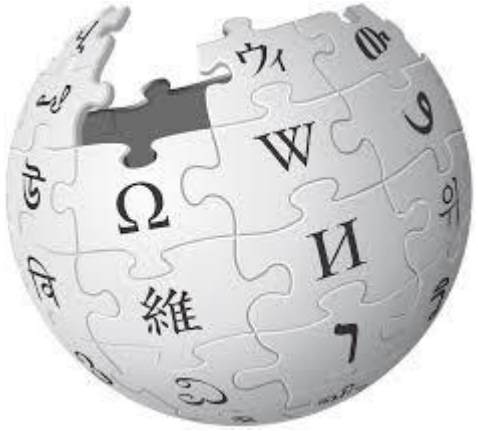
Encyclopedia

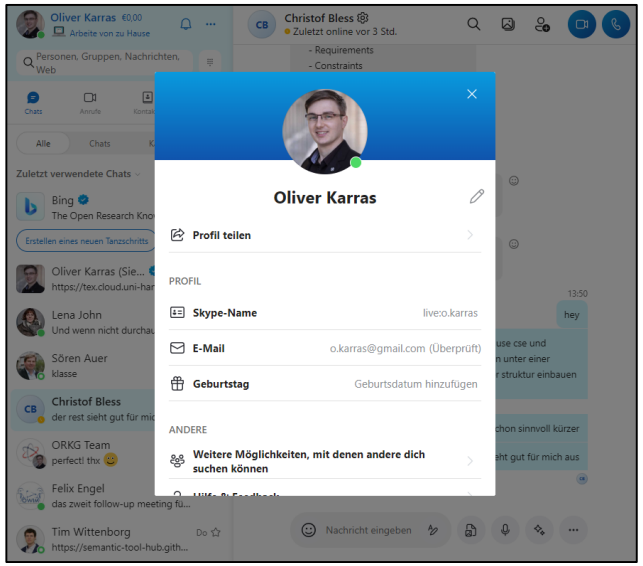
Phone books

… until Digital Transformation (Digitalization)!









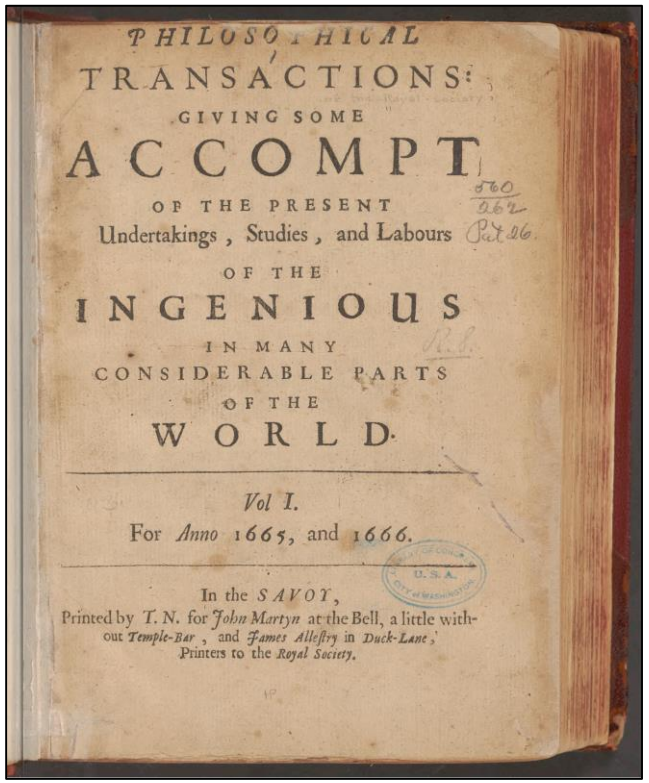
The World of Publishing & Communication has Profundely Changed!

- **New means adapted to the new possibilities**, e.g., platforms
- Completely new **business models**
- More focus on **data, interlinking, services,** and **search**
- **Integration**, **crowdsourcing, and data curation** are important

What has happened in academia in terms of scholarly publishing & communication?

Let's Take a Look

17th century

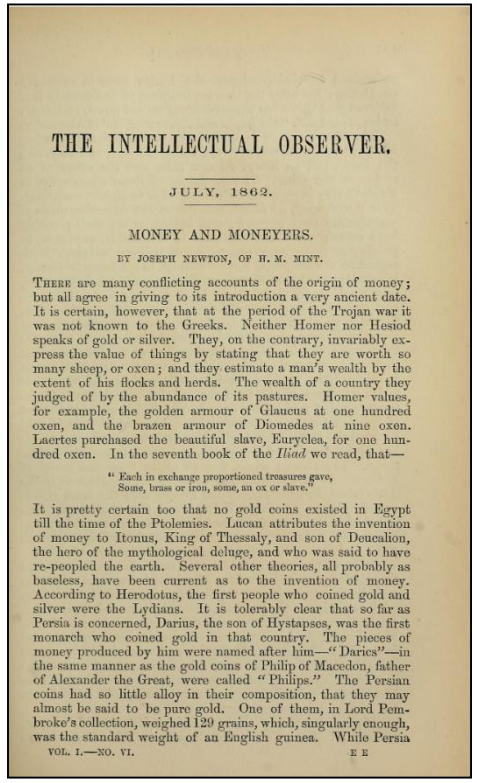


PHILOSOPHICAL
TRANSACTIONS:
GIVING SOME
ACCOMPT
OF THE PRESENT
Undertakings, Studies, and Labours
OF THE
INGENIOUS
IN MANY
CONSIDERABLE PARTS
OF THE
WORLD.

Vol I.
For *Anno* 1665, and 1666.

In the *SAVOY*,
Printed by *T. N.* for *John Martyn* at the Bell, a little without *Temple-Bar*, and *James Allestry* in *Duck-Lane*,
Printers to the *Royal Society*.

19th century



THE INTELLECTUAL OBSERVER.

JULY, 1862.

MONEY AND MONEYERS.

BY JOSEPH NEWTON, OF H. M. MINT.

THERE are many conflicting accounts of the origin of money; but all agree in giving to its introduction a very ancient date. It is certain, however, that at the period of the Trojan war it was not known to the Greeks. Neither Homer nor Hesiod speaks of gold or silver. They, on the contrary, invariably express the value of things by stating that they are worth so many sheep, or oxen; and they estimate a man's wealth by the extent of his flocks and herds. The wealth of a country they judged of by the abundance of its pastures. Homer values, for example, the golden armour of Glaucus at one hundred oxen, and the brazen armour of Diomedes at nine oxen. Laertes purchased the beautiful slave, Eurycles, for one hundred oxen. In the seventh book of the *Iliad* we read, that—

"Each in exchange proportioned treasures gave,
Some, brass or iron, some, an ox or slave."

It is pretty certain too that no gold coins existed in Egypt till the time of the Ptolemies. Lucan attributes the invention of money to Itonus, King of Thessaly, and son of Deucalion, the hero of the mythological deluge, and who was said to have re-peopled the earth. Several other theories, all probably as baseless, have been current as to the invention of money. According to Herodotus, the first people who coined gold and silver were the Lydians. It is tolerably clear that so far as Persia is concerned, Darius, the son of Hystapses, was the first monarch who coined gold in that country. The pieces of money produced by him were named after him—"Darics"—in the same manner as the gold coins of Philip of Macedon, father of Alexander the Great, were called "Philips." The Persian coins had so little alloy in their composition, that they may almost be said to be pure gold. One of them, in Lord Pembroke's collection, weighed 129 grains, which, singularly enough, was the standard weight of an English guinea. While Persia

VOL. I.—NO. VI. E E

20th century

Information Retrieval P. BAXENDALE, Editor

A Relational Model of Data for Large Shared Data Banks

E. F. CODD
IBM Research Laboratory, San Jose, California

Future users of large data banks must be protected from having to know how the data is organized in the machine (the internal representation). A prompting service which supplies such information is not a satisfactory solution. Activities of users at terminals and most application programs should remain unaffected when the internal representation of data is changed and even when some aspects of the external representation are changed. Changes in data representation will often be needed as a result of changes in query, update, and report traffic and natural growth in the types of stored information.

Existing noninferential, formatted data systems provide users with tree-structured files or slightly more general network models of the data. In Section 1, inadequacies of these models are discussed. A model based on *n*-ary relations, a normal form for data base relations, and the concept of a universal data sublanguage are introduced. In Section 2, certain operations on relations (other than logical inference) are discussed and applied to the problems of redundancy and consistency in the user's model.

KEY WORDS AND PHRASES: data bank, data base, data structure, data organization, hierarchies of data, networks of data, relations, derivability, redundancy, consistency, composition, join, retrieval language, predicate calculus, security, data integrity
CR CATEGORIES: 3.70, 3.73, 3.75, 4.20, 4.22, 4.29

1. Relational Model and Normal Form

1.1. INTRODUCTION

This paper is concerned with the application of elementary relation theory to systems which provide shared access to large banks of formatted data. Except for a paper by Childs [1], the principal application of relations to data systems has been to deductive question-answering systems. Levein and Maron [2] provide numerous references to work in this area.

In contrast, the problems treated here are those of *data independence*—the independence of application programs and terminal activities from growth in data types and changes in data representation—and certain kinds of *data inconsistency* which are expected to become troublesome even in nondeductive systems.

The relational view (or model) of data described in Section 1 appears to be superior in several respects to the graph or network model [3, 4] presently in vogue for noninferential systems. It provides a means of describing data with its natural structure only—that is, without superimposing any additional structure for machine representation purposes. Accordingly, it provides a basis for a high level data language which will yield maximal independence between programs on the one hand and machine representation and organization of data on the other.

A further advantage of the relational view is that it forms a sound basis for treating derivability, redundancy, and consistency of relations—these are discussed in Section 2. The network model, on the other hand, has spawned a number of confusions, not the least of which is mistaking the derivation of connections for the derivation of relations (see remarks in Section 2 on the "connection trap").

Finally, the relational view permits a clearer evaluation of the scope and logical limitations of present formatted data systems, and also the relative merits (from a logical standpoint) of competing representations of data within a single system. Examples of this clearer perspective are cited in various parts of this paper. Implementations of systems to support the relational model are not discussed.

1.2. DATA DEPENDENCIES IN PRESENT SYSTEMS

The provision of data description tables in recently developed information systems represents a major advance toward the goal of data independence [5, 6, 7]. Such tables facilitate changing certain characteristics of the data representation stored in a data bank. However, the variety of data representation characteristics which can be changed *without logically impairing some application programs* is still quite limited. Further, the model of data with which users interact is still cluttered with representational properties, particularly in regard to the representation of collections of data (as opposed to individual items). Three of the principal kinds of data dependencies which still need to be removed are: ordering dependence, indexing dependence, and access path dependence. In some systems these dependencies are not clearly separable from one another.

1.2.1. *Ordering Dependence.* Elements of data in a data bank may be stored in a variety of ways, some involving no concern for ordering, some permitting each element to participate in one ordering only, others permitting each element to participate in several orderings. Let us consider those existing systems which either require or permit data elements to be stored in at least one total ordering which is closely associated with the hardware-determined ordering of addresses. For example, the records of a file concerning parts might be stored in ascending order by part serial number. Such systems normally permit application programs to assume that the order of presentation of records from such a file is identical to (or is a subordering of) the

21st century

AGDISTIS - Graph-Based Disambiguation of Named Entities Using Linked Data

Ricardo Usbeck[1,2], Axel-Cyrille Ngonga Ngomo[1], Michael Röder[1,2], Daniel Gerber[1], Sandro Athaide Coelho[3], Sören Auer[4], and Andreas Both[2]

[1] University of Leipzig, Germany
[2] R&D, Unister GmbH, Germany
[3] Federal University of Juiz de Fora, Brazil
[4] University of Bonn & Fraunhofer IAIS, Germany
{usbeck,ngonga}@informatik.uni-leipzig.de

Abstract. Over the last decades, several billion Web pages have been made available on the Web. The ongoing transition from the current Web of unstructured data to the Web of Data yet requires scalable and accurate approaches for the extraction of structured data in RDF (Resource Description Framework) from these websites. One of the key steps towards extracting RDF from text is the disambiguation of named entities. While several approaches aim to tackle this problem, they still achieve poor accuracy. We address this drawback by presenting AGDISTIS, a novel knowledge-base-agnostic approach for named entity disambiguation. Our approach combines the Hypertext-Induced Topic Search (HITS) algorithm with label expansion strategies and string similarity measures. Based on this combination, AGDISTIS can efficiently detect the correct URIs for a given set of named entities within an input text. We evaluate our approach on eight different datasets against state-of-the-art named entity disambiguation frameworks. Our results indicate that we outperform the state-of-the-art approach by up to 29% F-measure.

1 Introduction

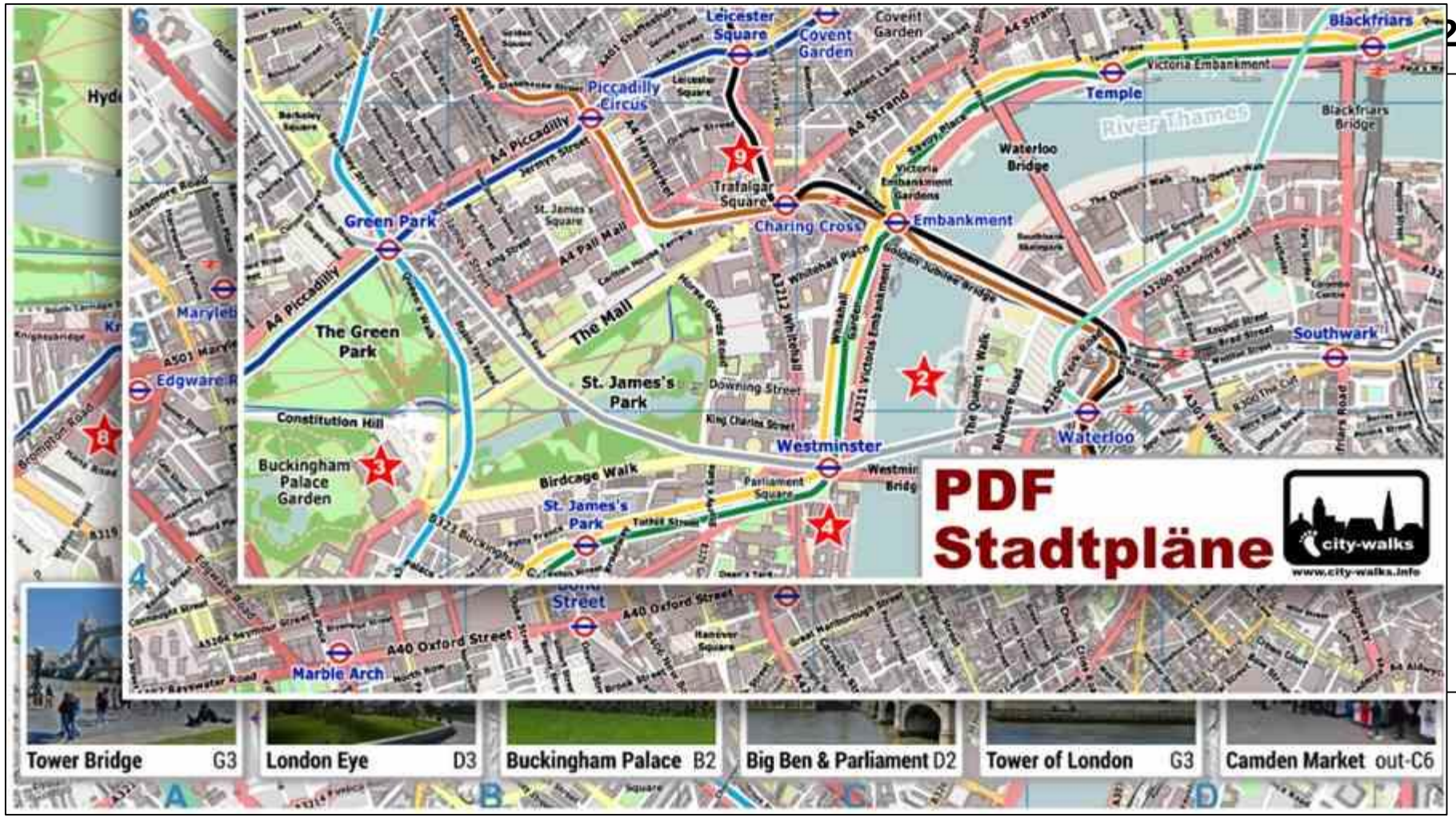
The vision behind the Web of Data is to provide a new machine-readable layer to the Web where the content of Web pages is annotated with structured data (e.g., RDFa [1]). However, the Web in its current form is made up of at least 15 billion Web pages.[1] Most of these websites are unstructured in nature. Realizing the vision of a usable and up-to-date Web of Data thus requires scalable and accurate natural-language-processing approaches that allow extracting RDF from such unstructured data. Three tasks play a central role when extracting RDF from unstructured data: named entity recognition (NER), named entity disambiguation (NED), also known as entity linking [16], and relation extraction (RE). For the first sentence of Example 1, an accurate named entity recognition approach would return the strings `Barack Obama` and `Washington`,

[1] Data gathered from `http://www.worldwidewebsize.com/` on January 4th, 2014.

P. Mika et al. (Eds.) ISWC 2014, Part I, LNCS 8796, pp. 457–471, 2014.
© Springer International Publishing Switzerland 2014

Scholarly publishing & communication has **not changed** (much)!

Let's Take a Look



21st century

AGDISTIS - Graph-Based Disambiguation of Named Entities Using Linked Data

Ricardo Usbeck[1,2], Axel-Cyrille Ngonga Ngomo[1], Michael Röder[1,2], Daniel Gerber[1], Sandro Athaide Coelho[3], Sören Auer[4], and Andreas Both[2]

[1] University of Leipzig, Germany
[2] R&D, Unister GmbH, Germany
[3] Federal University of Juiz de Fora, Brazil
[4] University of Bonn & Fraunhofer IAIS, Germany
{usbeck,ngonga}@informatik.uni-leipzig.de

Abstract. Over the last decades, several billion Web pages have been made available on the Web. The ongoing transition from the current Web of unstructured data to the Web of Data yet requires scalable and accurate approaches for the extraction of structured data in RDF (Resource Description Framework) from these websites. One of the key steps towards extracting RDF from text is the disambiguation of named entities. While several approaches aim to tackle this problem, they still achieve poor accuracy. We address this drawback by presenting AGDISTIS, a novel knowledge-base-agnostic approach for named entity disambiguation. Our approach combines the Hypertext-Induced Topic Search (HITS) algorithm with label expansion strategies and string similarity measures. Based on this combination, AGDISTIS can efficiently detect the correct URIs for a given set of named entities within an input text. We evaluate our approach on eight different datasets against state-of-the-art named entity disambiguation frameworks. Our results indicate that we outperform the state-of-the-art approach by up to 29% F-measure.

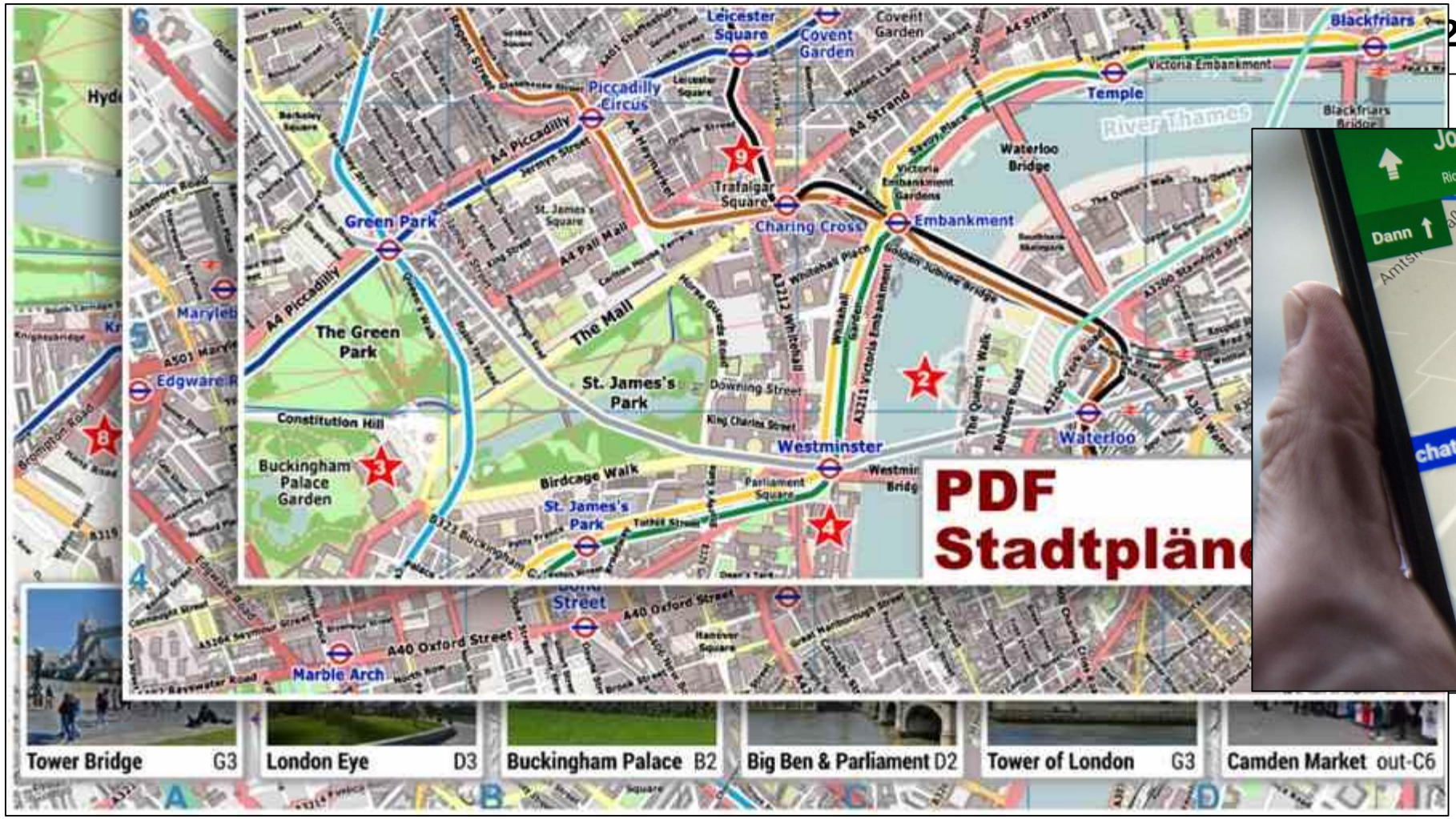
1 Introduction

The vision behind the Web of Data is to provide a new machine-readable layer to the Web where the content of Web pages is annotated with structured data (e.g., RDFa [1]). However, the Web in its current form is made up of at least 15 billion Web pages.[1] Most of these websites are unstructured in nature. Realizing the vision of a usable and up-to-date Web of Data thus requires scalable and accurate natural-language-processing approaches that allow extracting RDF from such unstructured data. Three tasks play a central role when extracting RDF from unstructured data: named entity recognition (NER), named entity disambiguation (NED), also known as entity linking [16], and relation extraction (RE). For the first sentence of Example 1, an accurate named entity recognition approach would return the strings `Barack Obama` and `Washington,`

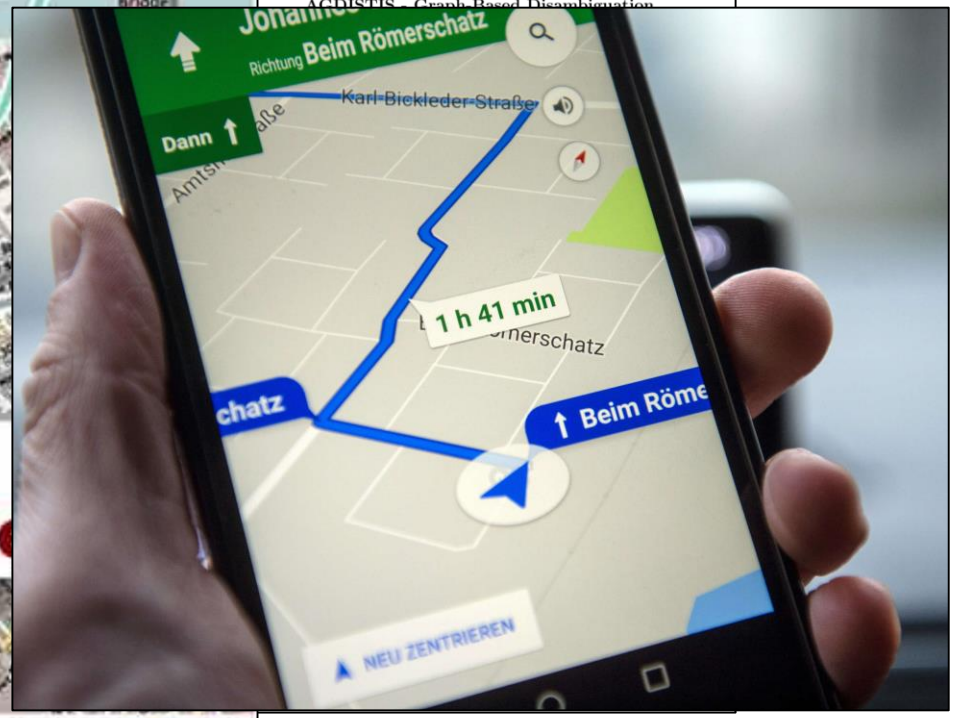
[1] Data gathered from `http://www.worldwidewebsize.com/` on January 4th, 2014.

P. Mika et al. (Eds.) ISWC 2014, Part I, LNCS 8796, pp. 457–471, 2014.
© Springer International Publishing Switzerland 2014

Let's Take a Look



21st century



Rethink How Scientific Knowledge is Communicated

"*The lightbulb was **not** invented by improving the candle.*"

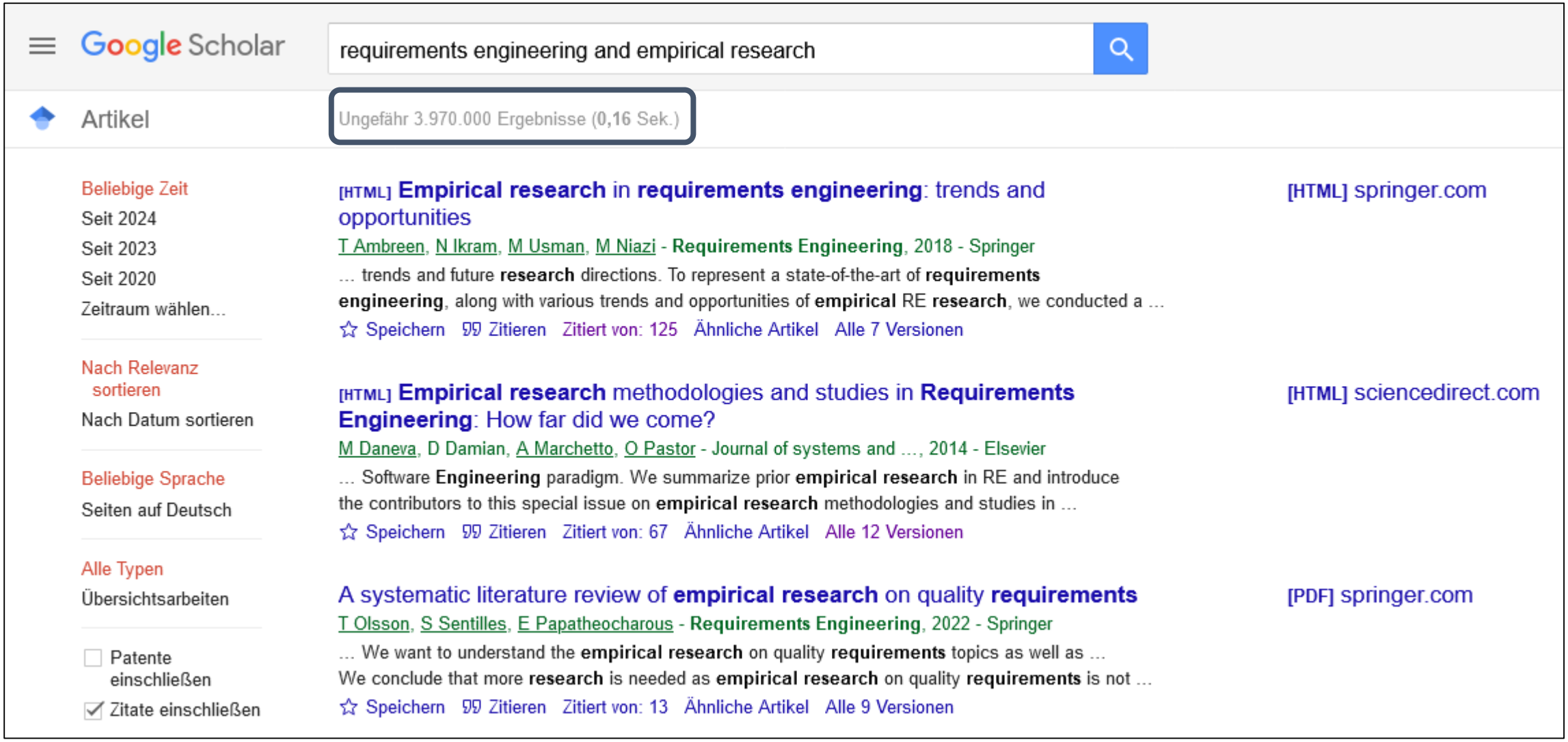
Oren Harari

Digitalization is **more** than just Digitization!

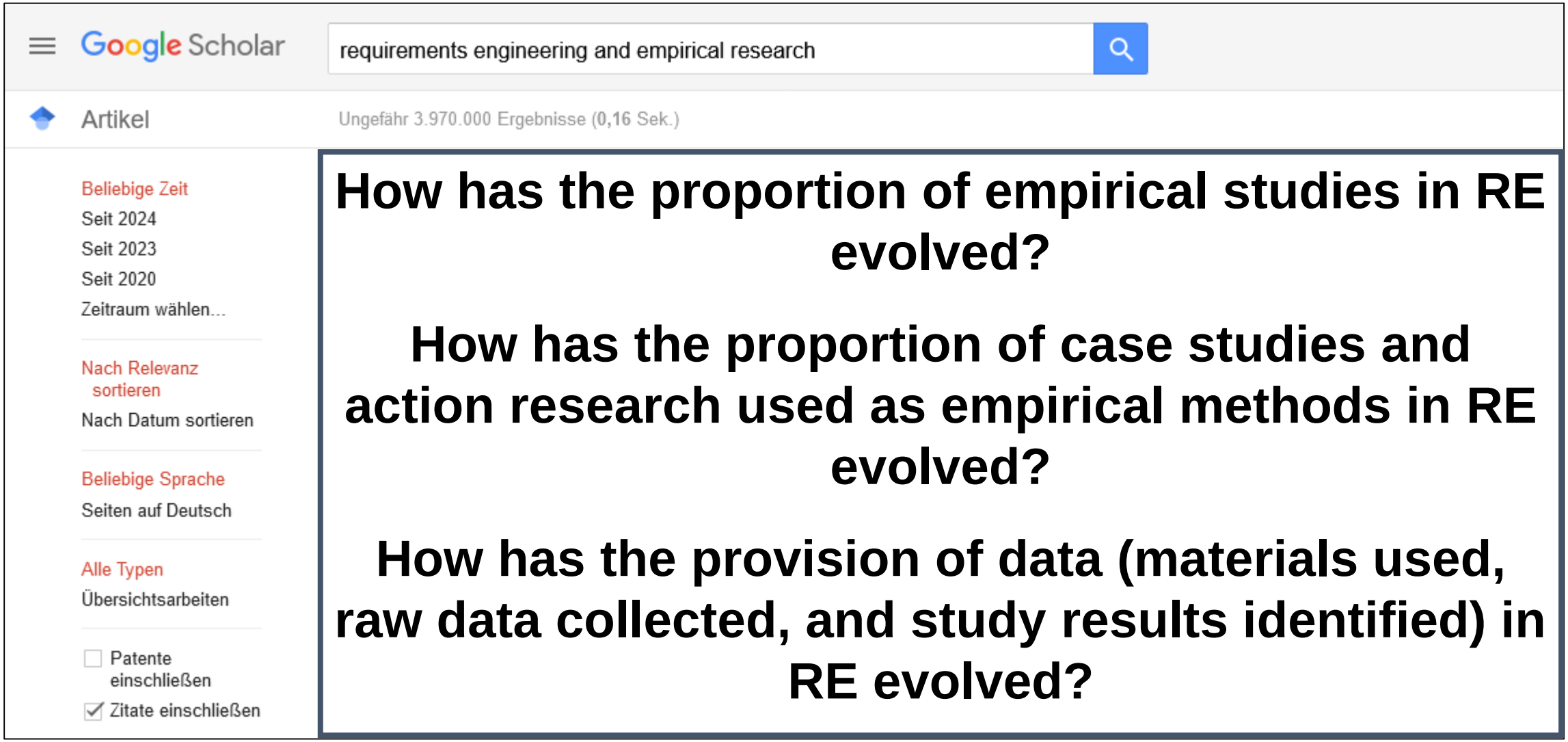
Current and future scientific challenges can not be tackled with an outdated communication system.

Digitalize Knowledge, Not Documents!

Example: Requirements Engineering and Empirical Research



Example: Requirements Engineering and Empirical Research



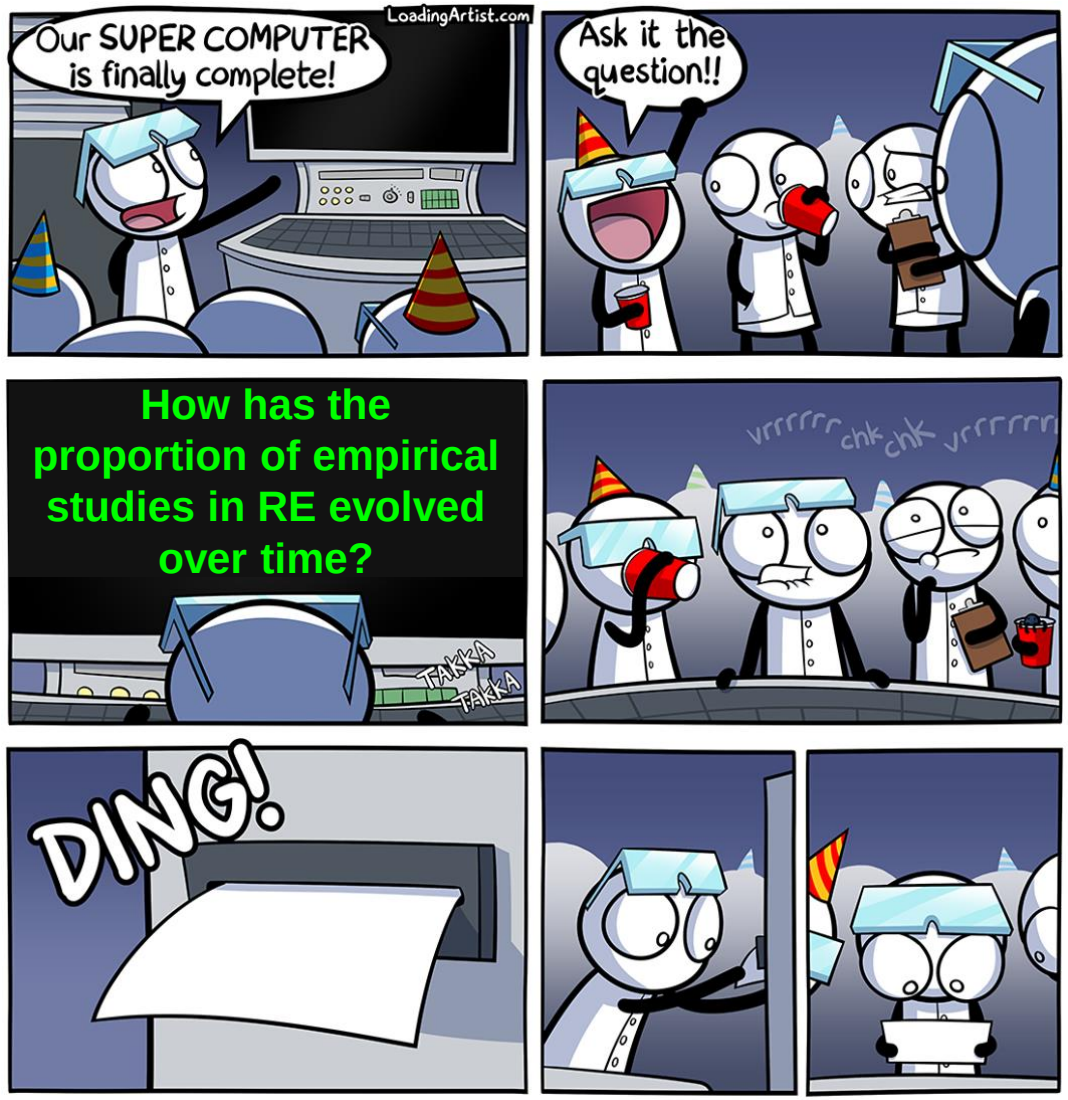
How do We Answer These Questions so far?

(S)**LR**s and SMSs for a **comprehensive**, **up-to-date**, and **long-term available** overview.

Over **7 million publications** per year[4] with an **increasing** (exponential) tendency.

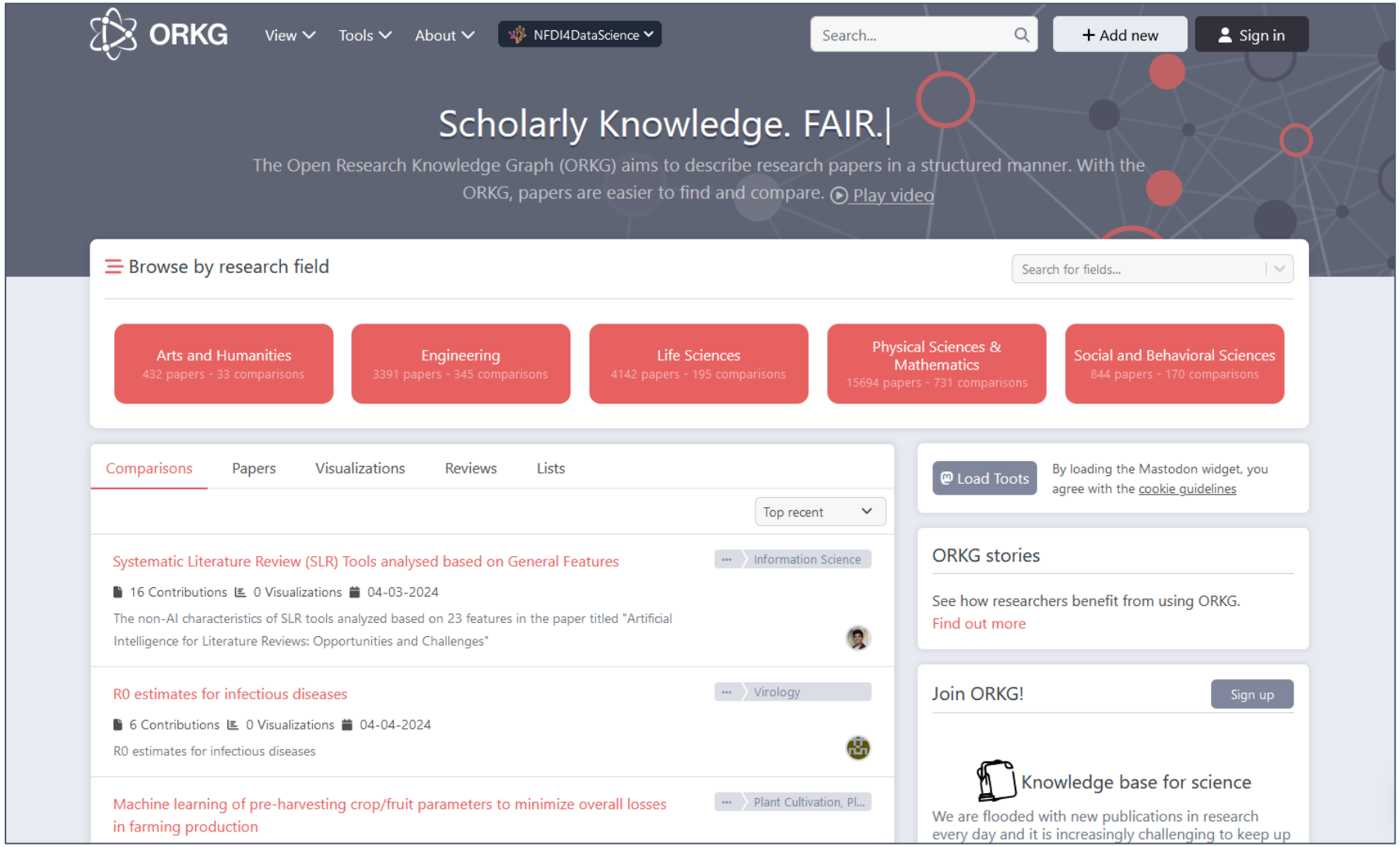
[4] Fire and Guestrin: *Over-Optimization of Academic Publishing Metrics: Observing Goodhart's Law in Action.* GigaScience Vol. 8, No. 6., DOI: 10.1093/gigascience/giz053, 2019.

Wouldn’t it be Great if we Could Ask the Computer?



How can we achieve this goal?

Open Research Knowledge Graph (ORKG)



The ORKG **revolutionizes** how **scientific knowledge** is communicated and processed, making it **actionable** for **machines** and **researchers** to navigate, compare, and review **vast amounts of information** efficiently.

Research Knowledge Graphs

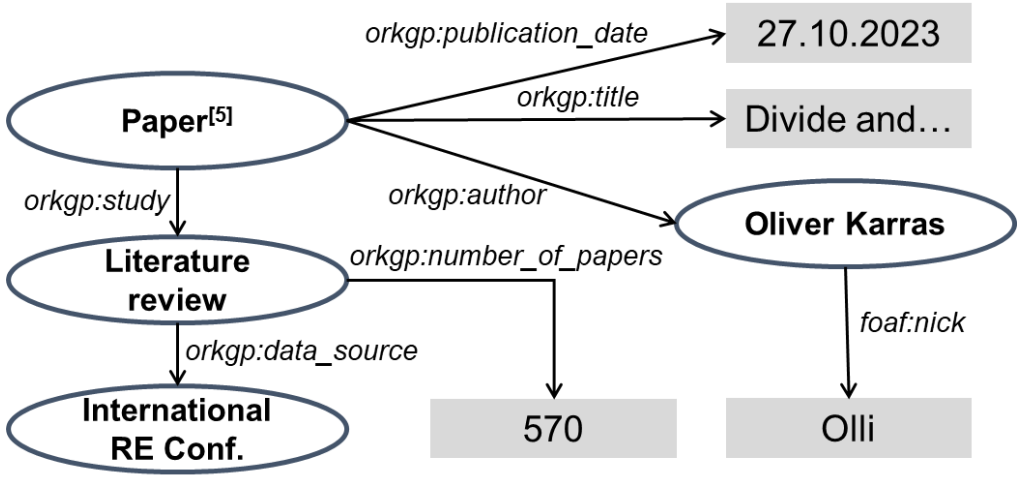
Research Knowledge Graphs (RKGs) are a **technology** for organizing scientific (meta-)data in a

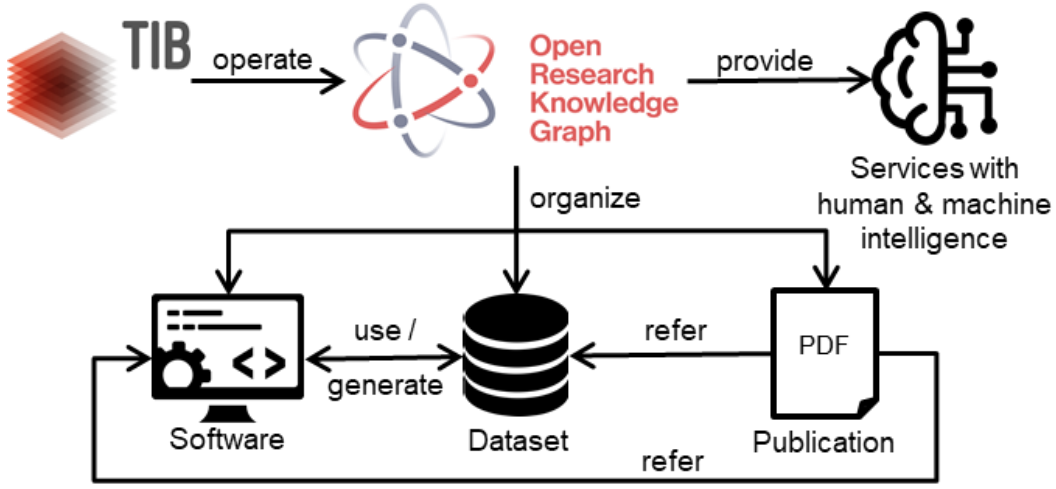
- **Flexible**, **fine-grained**, and **semantic** representation
- That is **understandable** and **processable** by humans and machines

Research Knowledge Graphs

Research Knowledge Graphs (RKGs) are a **technology** for organizing scientific (meta-)data in a

- **Flexible**, **fine-grained**, and **semantic** representation
- That is **understandable** and **processable** by humans and machines

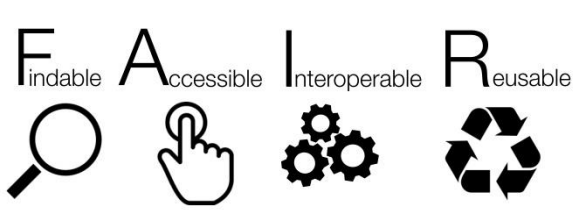




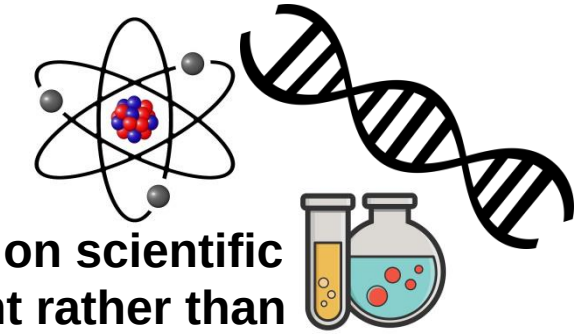
The **ORKG** is a **ready-to-use** and **sustainably** operated **infrastructure** with **services** that uses a **cross-discipline RKG** for the **long-term** and **openly available organization** of scientific (meta-)data according to the **FAIR** data principles.

[5] Karras et al.: *Divide and Conquer the EmpiRE: A Community-Maintainable Knowledge Graph of Empirical Research in Requirements Engineering.* 2023 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), DOI: 10.1109/ESEM56168.2023.10304795, 2023.

ORKG Objectives



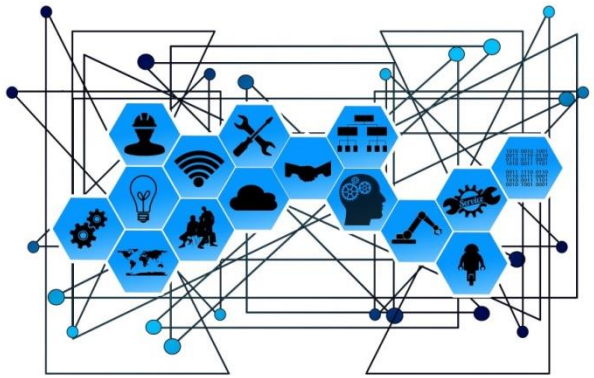
Make research FAIR



Focus on scientific content rather than pure metadata

Foster collaboration

Provide long-term available and up-to-date overviews



Tackle interdisciplinary challenges

Structured Representation of Scientific Knowledge

Mining User Requirements from Application Store Reviews Using Frame Semantics

Nishant Jha and Anas Mahmoud(✉)

The Division of Computer Science and Engineering, Louisiana State University, Baton Rouge, LA 70803, USA
njha1@lsu.edu, mahmoud@csc.lsu.edu

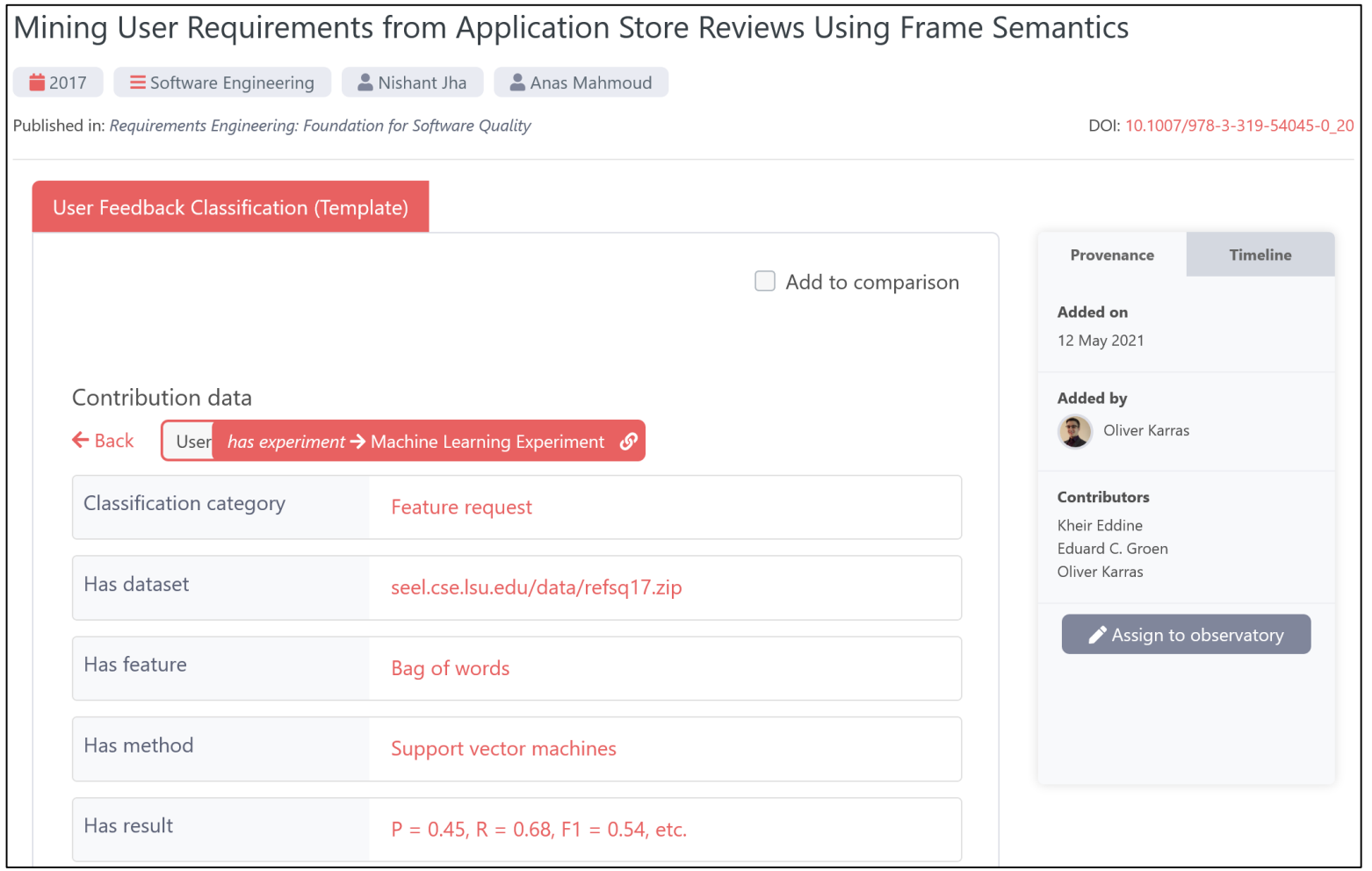
Abstract. *Context and motivation*: Research on mining user reviews in mobile application (app) stores has noticeably advanced in the past few years. The majority of the proposed techniques rely on classifying the textual description of user reviews into different categories of technically informative user requirements and uninformative feedback. *Question/Problem*: Relying on the textual attributes of reviews often produces high dimensional models. This increases the complexity of the classifier and can lead to overfitting problems. *Principal ideas/results*: We propose a novel semantic approach for app review classification. The proposed approach is based on the notion of semantic role labeling, or characterizing the lexical meaning of text in terms of semantic frames. Semantic frames help to generalize from text (individual words) to more abstract scenarios (contexts). This reduces the dimensionality of the data and enhances the predictive capabilities of the classifier. Three datasets of user reviews are used to conduct our experimental analysis. Results show that semantic frames can be used to generate lower dimensional and more accurate models in comparison to text classification methods. *Contribution*: A novel semantic approach for extracting user requirements from app reviews. The proposed approach enables a more efficient classification process and reduces the chance of overfitting.

Keywords: Requirements elicitation · Application stores · Classification

1 Introduction

Mobile application markets, or app stores (e.g., Google Play and Apple App Store), represent a unique model of service-oriented business. Such platforms have created an unprecedented opportunity for app developers to directly monitor the opinions of a large population of end-users of their software [25]. Through app stores feedback services, app users can directly share their experience in the form of textual reviews and meta-data (e.g., star ratings). Analyzing large datasets of app store reviews has revealed that they contain substantial amounts of up-to-date technical information. Such information can be leveraged by app developers to help them maintain and sustain their apps in a highly-competitive

© Springer International Publishing AG 2017
P. Grünbacher and A. Perini (Eds.): REFSQ 2017, LNCS 10153, pp. 273–287, 2017.
DOI: 10.1007/978-3-319-54045-0_20



Behind the Scenes

Mining User Requirements from Application Store Reviews Using Frame Semantics

Nishant Jha and Anas Mahmoud(✉)

The Division of Computer Science and Engineering, Louisiana State University, Baton Rouge, LA 70803, USA
njha1@lsu.edu, mahmoud@csc.lsu.edu

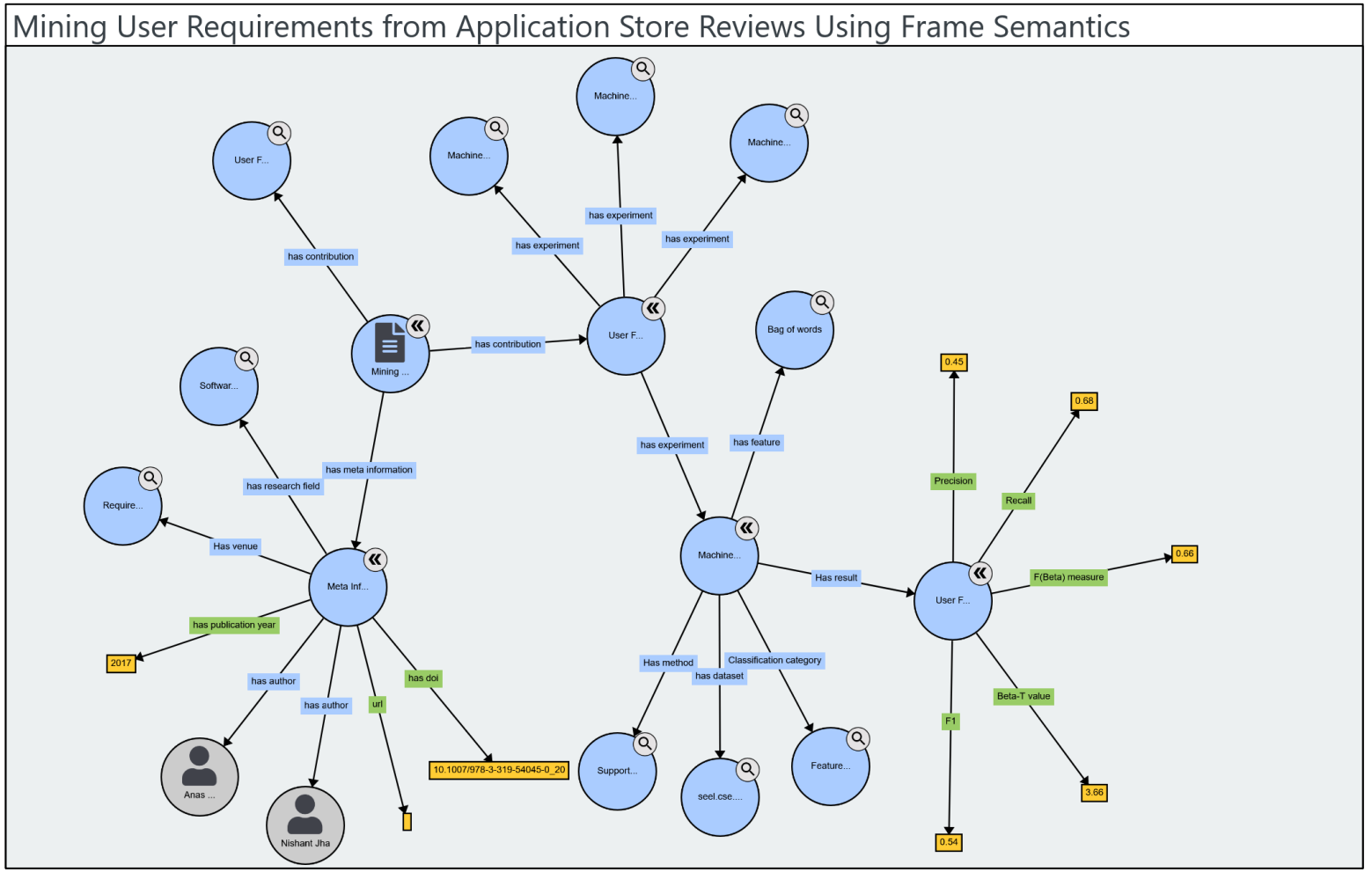
Abstract. *Context and motivation*: Research on mining user reviews in mobile application (app) stores has noticeably advanced in the past few years. The majority of the proposed techniques rely on classifying the textual description of user reviews into different categories of technically informative user requirements and uninformative feedback. *Question/Problem*: Relying on the textual attributes of reviews often produces high dimensional models. This increases the complexity of the classifier and can lead to overfitting problems. *Principal ideas/results*: We propose a novel semantic approach for app review classification. The proposed approach is based on the notion of semantic role labeling, or characterizing the lexical meaning of text in terms of semantic frames. Semantic frames help to generalize from text (individual words) to more abstract scenarios (contexts). This reduces the dimensionality of the data and enhances the predictive capabilities of the classifier. Three datasets of user reviews are used to conduct our experimental analysis. Results show that semantic frames can be used to generate lower dimensional and more accurate models in comparison to text classification methods. *Contribution*: A novel semantic approach for extracting user requirements from app reviews. The proposed approach enables a more efficient classification process and reduces the chance of overfitting.

Keywords: Requirements elicitation · Application stores · Classification

1 Introduction

Mobile application markets, or app stores (e.g., Google Play and Apple App Store), represent a unique model of service-oriented business. Such platforms have created an unprecedented opportunity for app developers to directly monitor the opinions of a large population of end-users of their software [25]. Through app stores feedback services, app users can directly share their experience in the form of textual reviews and meta-data (e.g., star ratings). Analyzing large datasets of app store reviews has revealed that they contain substantial amounts of up-to-date technical information. Such information can be leveraged by app developers to help them maintain and sustain their apps in a highly-competitive

© Springer International Publishing AG 2017
P. Grünbacher and A. Perini (Eds.): REFSQ 2017, LNCS 10153, pp. 273–287, 2017.
DOI: 10.1007/978-3-319-54045-0_20



Using FAIR Scientific Knowledge

Mining User Requirements from Application Store Reviews Using Frame Semantics

Nishant Jha and Anas Mahmoud(✉)

The Division of Computer Science and Engineering, Louisiana State University, Baton Rouge, LA 70803, USA
njha1@lsu.edu, mahmoud@csc.lsu.edu

Abstract. *Context and motivation*: Research on mining user reviews in mobile application (app) stores has noticeably advanced in the past few years. The majority of the proposed techniques rely on classifying the textual description of user reviews into different categories of technically informative user requirements and uninformative feedback. *Question/Problem*: Relying on the textual attributes of reviews often produces high dimensional models. This increases the complexity of the classifier and can lead to overfitting problems. *Principal ideas/results*: We propose a novel semantic approach for app review classification. The proposed approach is based on the notion of semantic role labeling, or characterizing the lexical meaning of text in terms of semantic frames. Semantic frames help to generalize from text (individual words) to more abstract scenarios (contexts). This reduces the dimensionality of the data and enhances the predictive capabilities of the classifier. Three datasets of user reviews are used to conduct our experimental analysis. Results show that semantic frames can be used to generate lower dimensional and more accurate models in comparison to text classification methods. *Contribution*: A novel semantic approach for extracting user requirements from app reviews. The proposed approach enables a more efficient classification process and reduces the chance of overfitting.

Keywords: Requirements elicitation · Application stores · Classification

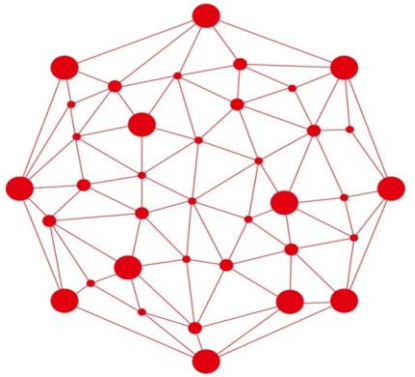
1 Introduction

Mobile application markets, or app stores (e.g., Google Play and Apple App Store), represent a unique model of service-oriented business. Such platforms have created an unprecedented opportunity for app developers to directly monitor the opinions of a large population of end-users of their software [25]. Through app stores feedback services, app users can directly share their experience in the form of textual reviews and meta-data (e.g., star ratings). Analyzing large datasets of app store reviews has revealed that they contain substantial amounts of up-to-date technical information. Such information can be leveraged by app developers to help them maintain and sustain their apps in a highly-competitive

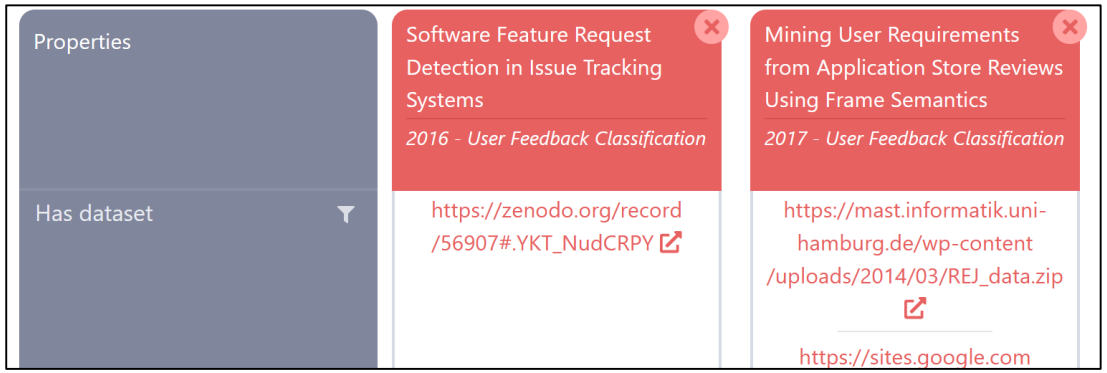
© Springer International Publishing AG 2017
P. Grünbacher and A. Perini (Eds.): REFSQ 2017, LNCS 10153, pp. 273–287, 2017.
DOI: 10.1007/978-3-319-54045-0_20

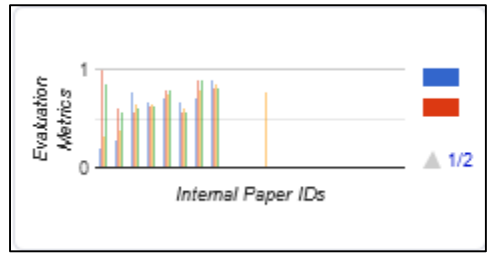
How has the provision of data (materials used, raw data collected, and study results identified) in RE evolved?

Natural language question



Answer





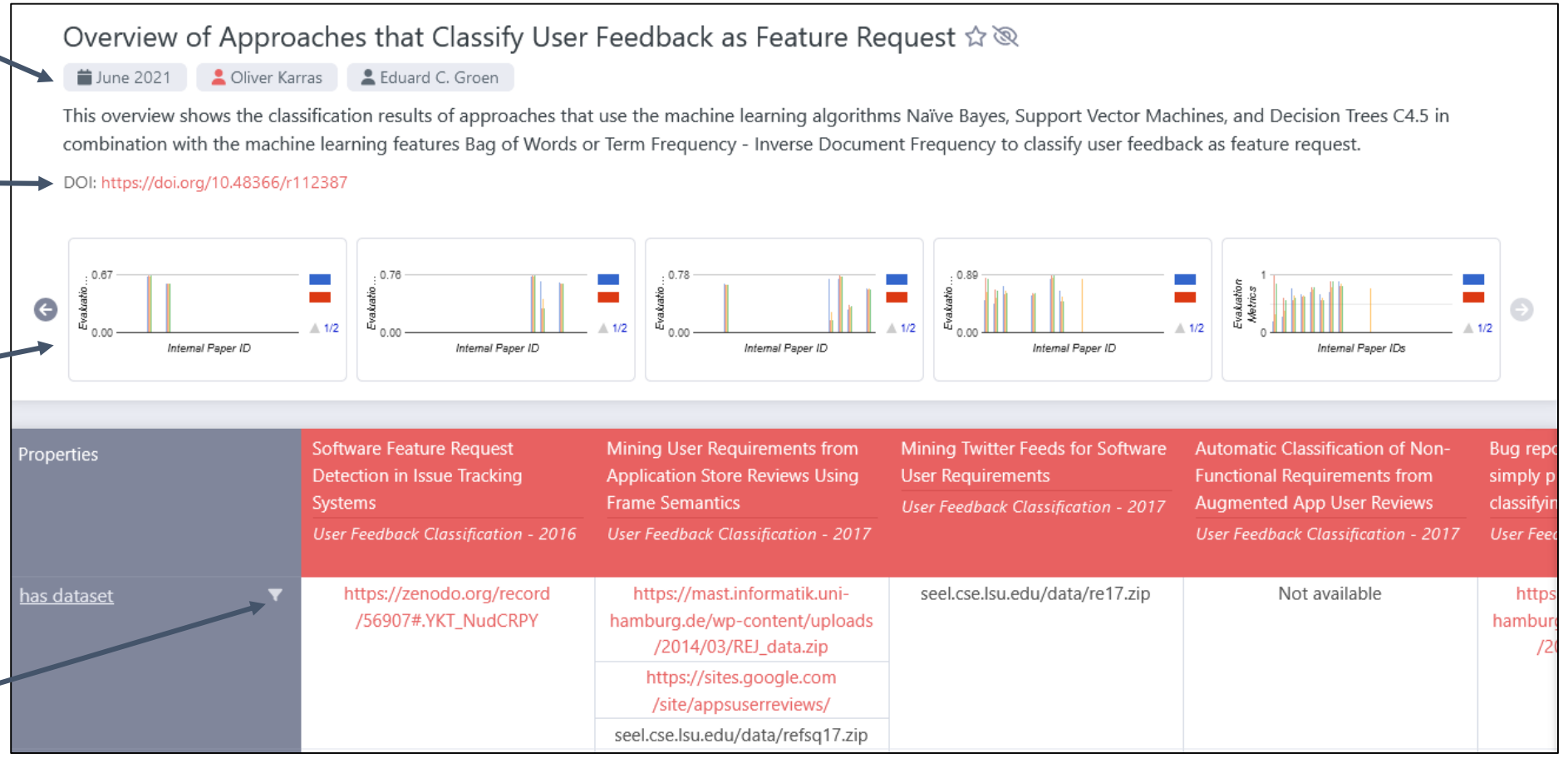
ORKG Comparisons

Acknowledgement of creators

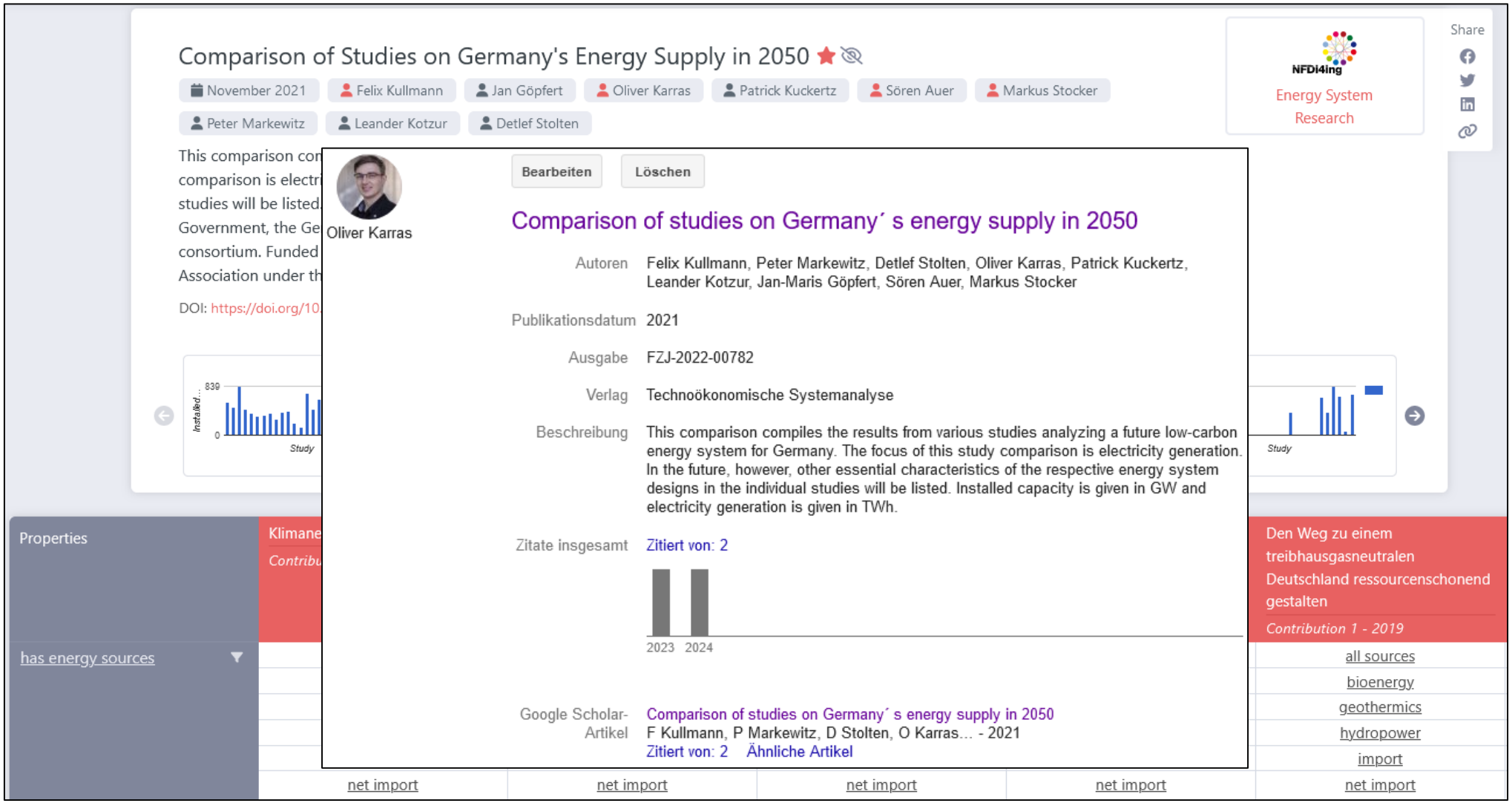
DOI

Visualizations

Interactive filtering



ORKG Comparisons are Citable



So far so Good, but…

…what can we do with machine-actionable scientific knowledge?

Anything we want!

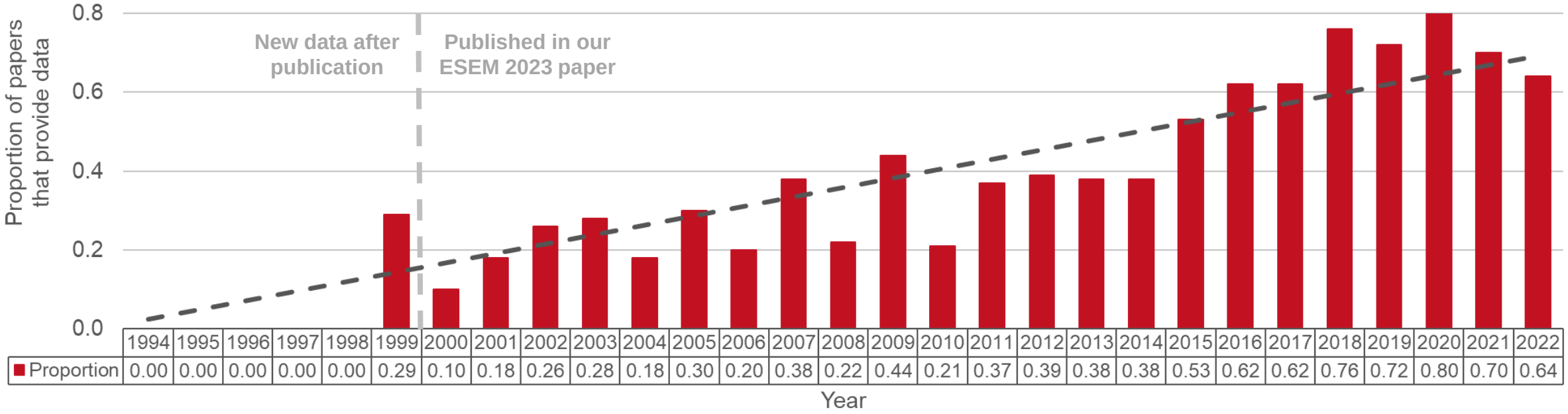
1. Papers, comparisons, and visualizations in the ORKG are **openly available for (re-)use and extension to anyone.**
2. The ORKG provides several **interfaces** for processing the data, e.g., to develop novel search, retrieval, mining, and assistance applications.

{ REST : API }

(Re-)use and Extension of Scientific Knowledge

How has the provision of data (materials used, raw data collected, and study results identified) in RE evolved?[5, 6, 7]

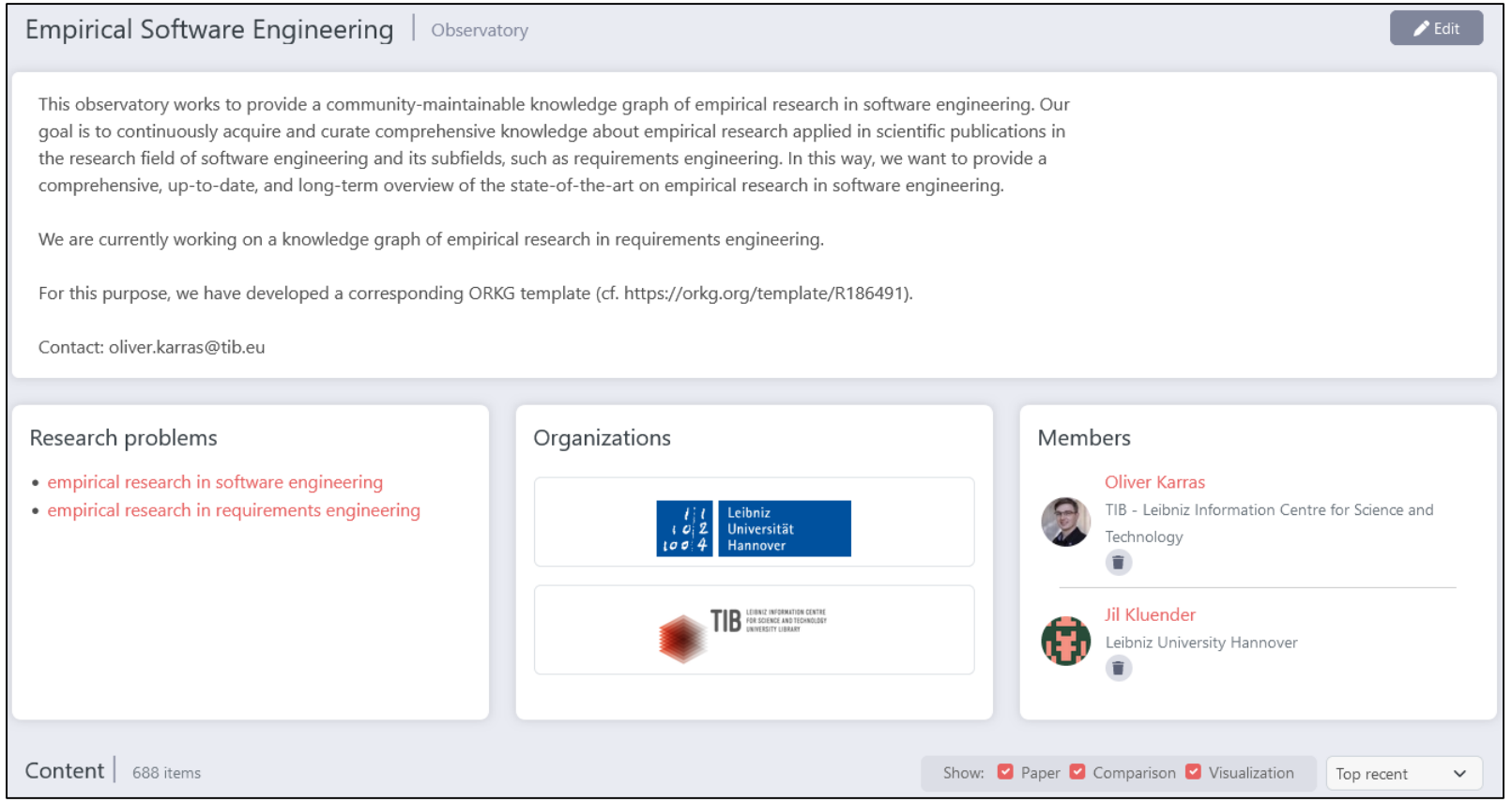
(Based on 680 papers in the ORKG)



	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Proportion	0.00	0.00	0.00	0.00	0.00	0.29	0.10	0.18	0.26	0.28	0.18	0.30	0.20	0.38	0.22	0.44	0.21	0.37	0.39	0.38	0.38	0.53	0.62	0.62	0.76	0.72	0.80	0.70	0.64

[5] Karras et al.: *Divide and Conquer the EmpiRE: A Community-Maintainable Knowledge Graph of Empirical Research in Requirements Engineering.* 2023 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), DOI: 10.1109/ESEM56168.2023.10304795, 2023.
[6] Project on GitHub: https://github.com/okarras/EmpiRE-Analysis
[7] Interactive Jupyter notebook: https://mybinder.org/v2/gh/okarras/EmpiRE-Analysis/HEAD?labpath=%2Fempire-analysis.ipynb

ORKG Observatory



- **Open** groups maintain topics in the ORKG
- **Central** access point for the community to all curated contents

Who organizes this scientific knowledge in the ORKG?

Who Organizes this Scientific Knowledge in the ORKG?

The heart of the ORKG is its **CROWD**:

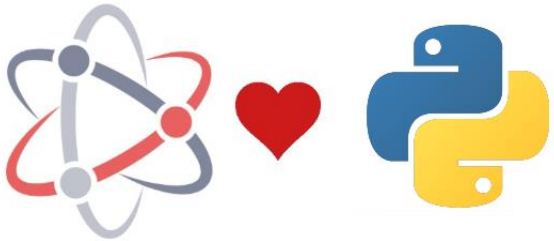
Researchers from any discipline!

How can We Contribute to the ORKG?

1. While writing a paper:
   - SciKGTeX: LaTeX package for FAIR annotations in a paper (embedded in the PDF) that can be imported into ORKG
   - https://github.com/Christof93/SciKGTeX

SciKGTeX

2. While developing an analysis:
   - Python package (& R package in development) for FAIR annotations in analysis scripts that can be imported into the ORKG
   - https://pypi.org/project/orkg/



3. At any time:
   - Manually using the ORKG Frontend to describe papers
   - (Semi-)automatically using the ORKG REST API

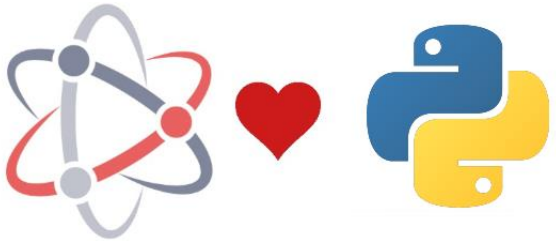
Open Research Knowledge Graph { REST : API }

What We Will Learn to Contribute to the ORKG

1. While writing a paper:
   - **SciKGTeX**: LaTeX package for FAIR annotations in a paper (embedded in the PDF) that can be imported into ORKG
   - https://github.com/Christof93/SciKGTeX

SciKGTeX

2. While developing an analysis:
   - Python package (& R package in development) for FAIR annotations in analysis scripts that can be imported into the ORKG
   - https://pypi.org/project/orkg/



3. At any time:
   - **Manually** using the ORKG Frontend to describe papers
   - (Semi-)automatically using the ORKG REST API

Open Research Knowledge Graph { REST : API }

Conclusion

Mining User Requirements from Application Store Reviews Using Frame Semantics

Nishant Jha and Anas Mahmoud(✉)

The Division of Computer Science and Engineering, Louisiana State University, Baton Rouge, LA 70803, USA
njha1@lsu.edu, mahmoud@csc.lsu.edu

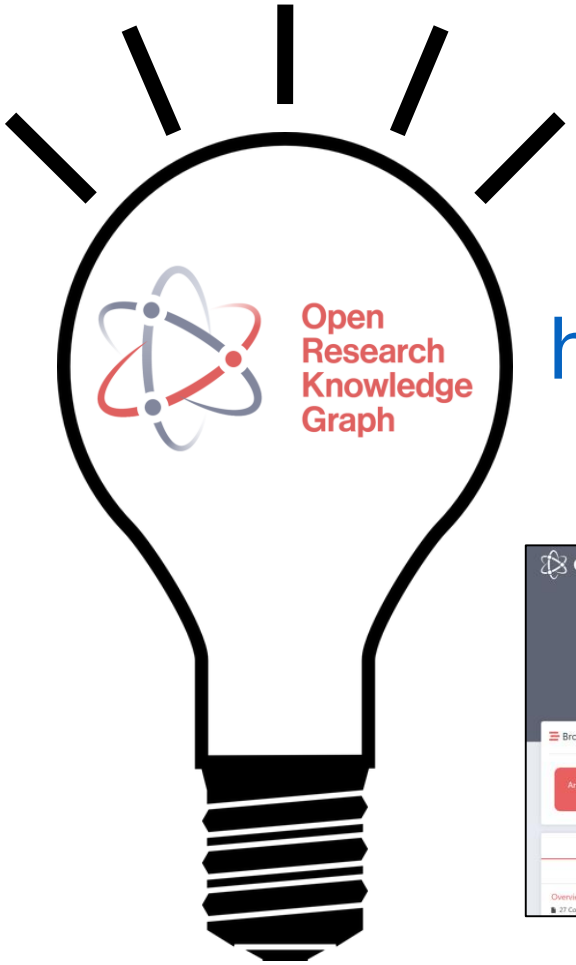
Abstract. *Context and motivation:* Research on mining user reviews in mobile application (app) stores has noticeably advanced in the past few years. The majority of the proposed techniques rely on classifying the textual description of user reviews into different categories of technically informative user requirements and uninformative feedback. *Question/Problem:* Relying on the textual attributes of reviews often produces high dimensional models. This increases the complexity of the classifier and can lead to overfitting problems. *Principal ideas/results:* We propose a novel semantic approach for app review classification. The proposed approach is based on the notion of semantic role labeling, or characterizing the lexical meaning of text in terms of semantic frames. Semantic frames help to generalize from text (individual words) to more abstract scenarios (contexts). This reduces the dimensionality of the data and enhances the predictive capabilities of the classifier. Three datasets of user reviews are used to conduct our experimental analysis. Results show that semantic frames can be used to generate lower dimensional and more accurate models in comparison to text classification methods. *Contribution:* A novel semantic approach for extracting user requirements from app reviews. The proposed approach enables a more efficient classification process and reduces the chance of overfitting.

Keywords: Requirements elicitation · Application stores · Classification

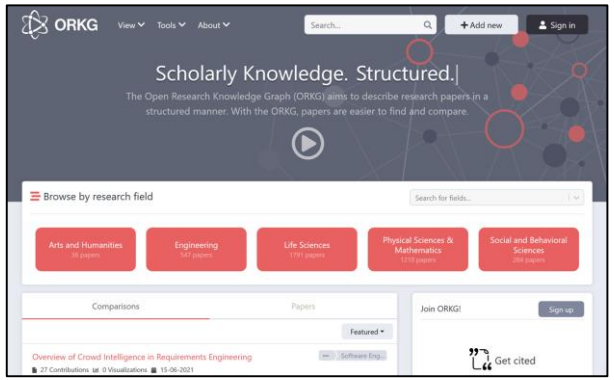
1 Introduction

Mobile application markets, or app stores (e.g., Google Play and Apple App Store), represent a unique model of service-oriented business. Such platforms have created an unprecedented opportunity for app developers to directly monitor the opinions of a large population of end-users of their software [25]. Through app stores feedback services, app users can directly share their experience in the form of textual reviews and meta-data (e.g., star ratings). Analyzing large datasets of app store reviews has revealed that they contain substantial amounts of up-to-date technical information. Such information can be leveraged by app developers to help them maintain and sustain their apps in a highly-competitive

© Springer International Publishing AG 2017
P. Grünbacher and A. Perini (Eds.): REFSQ 2017, LNCS 10153, pp. 273–287, 2017.
DOI: 10.1007/978-3-319-54045-0_20



https://orkg.org



Let's bring scholarly communication and open science in Requirements Engineering to the 21st century!

Further Reading

Open Access ORKG Book

- Celebrating the 5th anniversary of ORKG
- A practical guide for new and advanced users
  - ORKG's terms and concepts
  - ORKG's approach
  - ORKG's technology
  - ORKG's success stories

https://cuvillier.de/de/shop/publications/9037-open-research-knowledge-graph